

The Unintended Consequence of Diabetes Mellitus Pay-for-Performance (P4P) Program in Taiwan: Are Patients with More Comorbidities or More Severe Conditions Likely to Be Excluded from the P4P Program?

Tsung-Tai Chen, Kuo-Piao Chung, I-Chin Lin, and Mei-Shu Lai

Objective. Taiwan has instituted a pay-for-performance (P4P) program for diabetes mellitus (DM) patients that rewards doctors based in part on outcomes for their DM patients. Doctors are permitted to choose which of their DM patients are included in the P4P program. We test whether seriously ill DM patients are disproportionately excluded from the P4P program.

Data Source/Study Setting. This study utilizes data from the National Health Insurance (NHI) database in Taiwan for the period of January 2007 to December 2007. Our sample includes 146,481 DM-P4P patients (16.56 percent of the total) and 737,971 non-DM-P4P patients.

Data Collection/Extraction Methods. We use logistic and multilevel models to estimate the effects of patient and hospital characteristics on P4P selection.

Principal Findings. The results show that older patients and patients with more comorbidities or more severe conditions are prone to be excluded from P4P programs.

Conclusions. We found that DM patients are disproportionately excluded from P4P programs. Our results point to the importance of mandated participation and risk adjustment measures in P4P programs.

Key Words. Multilevel model, severity and comorbidity, unintended consequences, diabetes mellitus

Pay-for-performance (P4P) programs that reward providers based on outcome-based performance measures, or other “external” incentives that are not determined solely by provider behavior, can produce unintended consequences (Epstein, Lee, and Hamel 2004). This is important especially

when performance measures do not include risk adjustments to account for patient comorbidity or the severity of patients' conditions (Shen 2003). Providers who treat patients with more severe conditions may worry about unfair penalization because these patients are likely to cause a drop in providers' performance scores (Werner and Asch 2005). Incentives sometimes exist, therefore, to reap greater rewards by inappropriately excluding patients with more severe conditions. The potential for gaming the system using such adverse selection is problematic whenever providers are allowed to select patients for their P4P programs.

Of course, P4P programs could exclude patients appropriately in a number of circumstances. Programs might, for example, benefit from systematically excluding patients with characteristics that make them inappropriate for the measurement tools being used, or who require unique treatments (British Medical Association [BMA] 2009; Centre for Studies in Social Sciences 2009). However, this kind of exclusion (active exclusions) such as "exception reporting" may function also as acts of adverse selection.

One study showed little evidence of adverse selection, with hospitals reporting low rates of patient exclusion in P4P programs (Doran et al. 2008). Other studies reached the opposite conclusion, showing that adverse selection does indeed pose a significant problem (Doran et al. 2006; Sigfrid et al. 2006; Gravelle, Sutton, and Ma 2010). These latter three U.K. studies had a different context than our study generally. For instance, providers in the United Kingdom were not allowed to select the patients included in the P4P program, and these studies focused mostly on the exclusion rate for each hospital, not patient comorbidities or complications (Ryan 2009). A key objective in this paper is to evaluate the exclusion from P4P programs for diabetes mellitus (DM) of patients who have comorbidities or severe conditions. In this study, we hypothesize that providers are likely to exclude older patients and patients with high comorbidities or more severe conditions from P4P programs.

Address correspondence to Mei-Shu Lai, M.D. Ph.D., Room 518, 5F., No. 17, Syujhou Rd., Jhongjheng District, Taipei City, Taiwan, R.O.C.; e-mail: mslai@ntu.edu.tw. Tsung-Tai Chen, Ph.D., is with the Department of Healthcare Information and Management, Ming Chuan University and the Center for Health Insurance Research, College of Public Health, National Taiwan University. Kuo-Piao Chung, Ph.D., is with the Center for Health Insurance Research, College of Public Health, National Taiwan University and the Institute of Health Care Organization Administration, College of Public Health, National Taiwan University. I-Chin Lin, M.Sc., is with the Institute of Preventive Medicine, College of Public Health, National Taiwan University. Mei-Shu Lai, M.D. Ph.D., is with the Center for Health Insurance Research, College of Public Health, National Taiwan University and the Institute of Preventive Medicine, College of Public Health, National Taiwan University.

MATERIALS AND METHODS

Study Setting

The DM-P4P program designed by the National Health Insurance (NHI) in Taiwan has been the most comprehensive, mature P4P program in Taiwan. The program is voluntary, and it has had roughly two periods of evolution since 2001. In the first period, the program not only required health care providers to participate in clinical training to become certified in Taiwan's Diabetes Shared Care System (Chiou et al. 2001), but it also encouraged health care providers to increase monitoring and follow-up care for patients. In addition, providers were given the freedom to decide which patients were enrolled in their P4P programs. Before the end of 2006, financial incentives were given only for process-based services (e.g., hemoglobin A1C testing). Both "increased physician fees" and "case management fees" were provided in addition to regular fee-for-service reimbursements (Lee et al. 2010). The design in this period focused on rewarding process-based services, a feature that lasted into the next period.

In the second period of the program's evolution, initiated toward the end of 2006, the NHI started paying extra bonuses for treatment outcome measures. In particular, they included the presence of two poor outcomes as an indication of poor care, that is, providers were not rewarded for patients with poor outcomes because these were patients who do not receive better care; the two poor outcomes used were the "Percentage of A1C \geq 9.5 percent" and "Percentage of low-density lipoprotein (LDL) \geq 130 mg/dl." Importantly, these measures do not include risk adjustments.

Design in this period of the program's evolution focused not only on rewarding process services but also on rewarding intermediate outcome measures. In addition, the incentive structure for intermediate outcome measures adopted a so-called quality tournament, in which only the highest-performing 25 percent of providers received rewards. The rewards were paid according to a composite score. The calculation of the composite score is similar to that of the indicator average (Reeves et al. 2007). However, the DM-P4P program in Taiwan adopts the rank, and not the rate, of each indicator. For example, if there are two intermediate outcome indicators, then the indicator average for a provider is calculated by mean rank of the providers according to the two indicators. This score represents the mean rank at which each measure was met.

Our study was based on this current design that emphasizes intermediate outcomes and uses a quality tournament to disburse rewards.

Data Sources

Our data come from two secondary databases. One database contained information collected from the regular NHI claim data for the period from January 2007 to December 2007 and was used to obtain patient and hospital characteristics. The other database, the P4P database, was intended to supplement regular claims data. DM patient outcome data, such as A1C or LDL values, were reported by the hospitals themselves and were entered into the P4P-specific database automatically.

Definition of the Study Groups, the Plurality Algorithm

One Taiwanese study found using a survey questionnaire that the accuracy of diabetes diagnoses in the Taiwan NHI database was only 74.6 percent. Because of this failure in accuracy, our study required patients who were selected to have received a diagnosis of diabetes (ICD-9-CM 250) and to have undergone more than four outpatient visits. This selection process ensures that the accuracy of diagnosis achieved is 99.16 times greater than that for patients with ≤ 1 outpatient visit (Lin et al. 2005). We looked at all DM patients in the regular NHI claim data who were strictly defined by these criteria in 2007; we divided them into two groups for comparison (P4P patients versus non-P4P patients). After applying the plurality algorithm (described below), the total number of DM patients was 884,452. There were two essential criteria for identifying members of the P4P patient group. First, outcome data for the diabetic patients had to be available in the P4P database. Second, the patients had to have at least one “P14 \times ” code (internal code) in the regular claims database. Ultimately, the total number of DM-P4P patients in our sample was 146,481, and the total number of non-DM-P4P patients was 737,971. Although the physician participation rate was around 47 percent (10,720/22,952), only about 16.56 percent (146,481/884,452) of DM patients were eligible to be selected by the NHI P4P program. This is because many of the physicians who saw large numbers of DM patients did not participate.

To address the problem of patients' multiple visits to different hospitals and provider accountability, the assignment algorithm called the plurality provider algorithm was applied in our study (Pham et al. 2007). The algorithm assigns a patient to the physician (or practice) who billed for the greatest number of care visits in a given year. Ties between physicians were resolved by favoring the physician with the greatest total charges for that patient (Pham et al. 2007).

Independent Variables

The independent variables affecting the likelihood of a patient being enrolled in a DM-P4P program include age, gender, comorbidity (Meduru et al. 2007), severity/complication (Selby et al. 2001; Rosenzweig et al. 2002), and number of visits. These variables characterize the patient. Hospital characteristics (Doran et al. 2006; Doran et al. 2008) include patient volume, summary/baseline score of hospital in prior year, and hospital level. Definitions for some of these independent variables are discussed below.

We adopt the chronic illness with complexity (CIC) method for adjusting comorbidity in patients with multiple chronic diseases. This index includes nondiabetes physical illness complexity (e.g., any cancer), diabetes-related complexity, and mental illness and substance abuse complexity. We ignore diabetes-related complexity because it contains only three kinds of diabetes complications. Instead, we use the diabetes complications severity index (DCSI) and calculate the comorbidity count using CIC. For patient severity/complication, we adopted Selby and colleagues and Rosenzweig and colleagues's DCSI (Rosenzweig et al. 2002; Young et al. 2008), which includes seven categorizations of complications: retinopathy, nephropathy, neuropathy, cerebrovascular complications, cardiovascular complications, peripheral vascular disease, and metabolic complications. Finally, the number of visits for each patient (individual variable) and the P4P patient volume at each hospital (hospital variable) are counted after assigning patients to a specific physician according to the plurality algorithm. The other variable, called the baseline score, is derived from the Raw Sum Score in the previous year for the hospital (Reeves et al. 2007). The Raw Sum Score method is recommended by the Centers for Medicare and Medicaid Services (CMS) for aggregating indicators within conditions; in this method, the process for calculating the Raw Sum Score is to sum the numerators, sum the denominators, and then calculate the ratio of summed numerators to summed denominators (Shwartz et al. 2008). We used two dichotomous measures (A1C < 9.5 percent and LDL < 130 mg/dl) and two process measures (annual A1C and LDL tests) to construct the Raw Sum Score (baseline score). This baseline score captures what the CIC and DCSI measures do not capture because hospitals with a lower baseline score in the previous year were more likely to exclude patients in the current year under some conditions (Doran et al. 2008).

Statistical Model

We estimate logistical models to study how patient inclusion in the P4P program is related to hospital factors and patient characteristics. The dependent

variable in this model is a dummy variable taking the value 1 for patients excluded from the P4P program and 0 for those included. We analyzed the data using *SAS*, version 9.1. The effect of hospital factors and patient characteristics are multilevel issues (Young 2008). We also estimated hierarchical models using *HLM 6* to investigate the sensitivity of our logistic model results. Results are quite similar. The interpretation of the hierarchical model is available in the Appendix to this paper.

RESULTS

Table 1 presents the distribution of patient characteristics between the study group (P4P enrollment) and the reference group (no P4P enrollment). Among the patients who were enrolled in P4P, 51.92 percent were female, 41.34 percent had DCSI scores > 0, and 34.50 percent had at least one additional medical condition (comorbidity). The average patient age was 61.9 years, and the average number of patient visits was 9.37. The group of patients enrolled in P4P had a lower average age and fewer patients with high comorbidity (CIC

Table 1: Characteristics of DM-P4P and Non-DM-P4P Cases in 2007

	<i>P4P Patients (N= 146,481)</i>	<i>Non-P4P Patients (N= 737,971)</i>	
	<i>Mean (SD) or n (%)</i>	<i>Mean (SD) or n (%)</i>	<i>p-Value</i>
Age	61.90 (12.08)	63.24 (12.73)	< .001
Gender			< .001
Female	76,049 (51.92)	367,661 (49.82)	
Male	70,432 (48.08)	370,310 (50.18)	
DCSI score (DM severity)			< .001
0	85,927 (58.66)	455,137 (61.67)	
1	35,062 (23.94)	136,631 (18.51)	
2	17,081 (11.66)	100,605 (13.63)	
3	5,481 (3.74)	26,073 (3.53)	
4	2,072 (1.41)	13,995 (1.90)	
5+	858 (0.59)	5,530 (0.75)	
CIC count (DM comorbidity)			< .001
0	95,952 (65.50)	479,760 (65.01)	
1	36,412 (24.86)	183,283 (24.84)	
2	10,150 (6.93)	54,060 (7.33)	
3	2,768 (1.89)	14,878 (2.02)	
4+	1,199 (0.82)	5,990 (0.81)	
Number of visits	9.37 (4.75)	8.61 (4.90)	< .001

CIC, chronic illness with complexity; DCSI, diabetes complications severity index; DM, diabetes mellitus; P4P, pay for performance; SD, standard deviation.

count, χ^2 -test, $p < .001$) or severe conditions (DCSI score, χ^2 -test, $p < .001$) compared with the non-P4P group. The P4P group also had a higher number of visits (t -test, $p < .001$) and a greater proportion of females (χ^2 -test, $p < .001$).

We estimate a logistic model to determine whether patients with higher comorbidity or severity are prone to be excluded from P4P programs by hospitals. We include patient and hospital characteristics in the regression equation for the logistic model (Table 2). The variables of age, DCSI score, CIC count, and hospital level have a significant odds ratio (OR) for participating in P4P. The results show that older patients and patients with higher comorbidity or severity are prone to be excluded from P4P programs. As the DCSI score and CIC count increases from zero to four or five, the probability of being excluded also increases. Hospitals with lower baseline scores in the previous year (2006) are more likely to exclude patients from P4P programs in the current year (2007) (OR = 0.98, $p < .001$). Hospital size has a significant negative effect on the number of patients who participate in P4P programs. As hospital size increased, so did the probability that DM patients would be excluded from P4P programs as compared with the clinics. Discrimination in the logistic model is assessed using the C index, which is equivalent to the area under the receiver-operating characteristic curve. The model shows a high level of discrimination (0.72). The mixed-effects model shows similar results to those of the logistic model, with magnitudes that are also very similar. (See Table SA1 in the Appendix for a detailed interpretation and explanation.)

DISCUSSION

We found that patients with greater severity or comorbidity were more likely to be excluded from P4P programs no matter whether we used the logistic or mixed-effects models. In addition, we found that hospitals with a lower baseline score in the previous year (2006) were more likely to exclude patients in the current year (2007), perhaps because hospitals with lower baseline scores in the previous year may want to increase their benefits in the next year. This result is similar to the results of other related studies (Doran et al. 2008). Another finding is that our study, like the other studies, demonstrated that larger hospitals may be more likely to exclude patients from P4P programs (Doran et al. 2008). It is for this reason that 65 percent of the variance in patient participation is explained by hospital characteristics and only 35 percent by patient characteristics.

Table 2: Factors Associated with Exclusion of DM Patients from P4P Programs

	<i>Logistic Model OR (95% CI)</i>
Patient level	
Intercept	5.27 (5.23,5.30)***
Age (Ref: ≤ 63)	
> 63	1.32 (1.30,1.33)***
Gender (Ref: F)	
Male	1.11 (1.09,1.12)***
DCSI score (Ref: 0)	
1	0.81 (0.79,0.82)***
2	1.31 (1.29,1.34)***
3	1.12 (1.08,1.15)***
4	1.70 (1.62,1.79)***
5+	1.78 (1.64,1.92)***
CIC count (Ref: 0)	
1	1.08 (1.06,1.10)***
2	1.21 (1.18,1.24)***
3	1.12 (1.08,1.15)***
4+	1.33 (1.24,1.42)***
Number of visits	0.96 (0.95,0.96)***
Hospital level	
Level (Ref: clinic)	
Tertiary hospital	8.13 (7.90,8.36)***
Regional hospital	4.02 (3.94,4.10)***
District hospital	2.98 (2.92,3.05)***
Baseline score [†]	0.98 (0.98,0.98)***
Patient volume	1.00 (1.00,1.00)*
C index	0.72

* $p < .05$.

*** $p < .001$.

[†]Prior year.

CIC, chronic illness with complexity; DCSI, diabetes complications severity index; DM, diabetes mellitus; OR, odds ratio; P4P, pay for performance; Ref, reference group.

The primary policy implications of this study include a pronounced need to prevent adverse selections during P4P programs' patient selection process. Several approaches have been proposed for preventing adverse selection. The first is to set target thresholds below 100 percent. Physicians would earn the maximum financial reward without achieving the target for all patients (Fleetcroft et al. 2008). For example, the United Kingdom sets thresh-

olds of 40–50 percent for the measure $A1C \leq 7$ (BMA 2009). This approach is not suitable for Taiwan because it represents a design based in competition among hospitals (a dynamic threshold), not on a policy that would allow these hospitals to earn more money by exceeding the threshold (a fixed threshold).

A second approach would allow physicians to remove inappropriate patients from the calculation of quality achievement (exception reporting) (BMA 2009). Several authors (Doran et al. 2008; Fleetcroft et al. 2008) have observed that this approach offers three advantages: It is precise, can increase the acceptance rate of the P4P program because of its active exclusion design, and may also help to eliminate situations in which patients are refused care because of severe medical conditions. However, evidence continues to indicate that the benefits from this design may be hampered by abuses of the system (Doran et al. 2006; Sigfrid et al. 2006; Gravelle et al. 2010).

The final approach is to make risk adjustment for outcome or process measures (Landon et al. 2003; Asch et al. 2006). For payers, an appropriate risk adjustment framework is important because physicians' behavior can only be changed by incentives when they consider the data to be complete and accurate and the score calculation to be fair; otherwise there will be a backlash against the system or physicians may try to game the system (Bokhour et al. 2006). As noted by Ryan and colleagues, "in the absence of complete risk adjustment, providers may engage in statistical discrimination: the application of perceived group characteristics to individuals" (Ryan 2009). Statistical discrimination may make providers avoid patients on the basis of unmeasured severity.

America and the United Kingdom face problems of politics in the implementation of P4P programs (Gulland 2003a, b; Tanenbaum 2009). Taiwan's reforms face similar difficulties. In Taiwan, the design of the original 2001 P4P program had itself faced the problem of interest group politics (Chang 2004). To resolve opposition to the P4P program and to encourage providers to participate, the original design was compromised in a manner that allowed voluntary provider participation and free patient program enrollment. In addition, there was no time schedule to complete the evolution of the P4P program. In May of 2009, the latest version of the DM-P4P program achieved some reforms by establishing the requirement that providers reach the new P4P patient enrollment rate of 30 percent and by requiring that providers attain a volume of P4P patients greater than the mandated threshold (50 or more).

Several other factors contribute to the problems of implementation. The insufficient funds for implementation of the P4P program represent yet

another difficulty. The yearly DM-P4P cost is only about 3–5 percent of the total expenditure in DM care in Taiwan (Lai et al. 2009). The limited investment by the NHI makes it difficult to learn more about implementing mandated participation from programs in the United Kingdom because it may not cover the additional expenses required for the implementation of the P4P program by hospitals (Epstein 2006), which involves procedural changes such as the reporting of clinical data (Halladay et al. 2009).

Universal electronic health record data made it possible for the United Kingdom to measure the process and outcome for all DM patients, and exclusions were selected only as an active process (Curcin et al. 2010). Hence, while patient complexity may result in the exclusion, we cannot rule out the cost of reporting clinical data as an additional motivator for the decision to exclude patients. At the very least, this study proved that there is an association between patient complexity and exclusion from the P4P program.

There were some limitations in our study. First, our data are limited by having no recorded reasons for exclusions, so we cannot determine the specific reasons why patients were excluded from P4P programs. Second, because of considerations about patient privacy, we cannot link our claims data to the “cause of death” data file supported by the Department of Health. We can only calculate all-cause inpatient mortality, making it likely that the total number of deaths was underestimated in this study. However, Table SA2 in the Appendix shows that the exclusion of deaths affected the result only slightly using the logistic model. Third, while our interpretation and motivation are connected to physicians’ individual practices of inclusion and exclusion, our analysis occurs at the hospital level rather than at the physician level.

There are several reasons that we perform our analysis at the hospital level. First, although in Taiwan incentive calculations were oriented toward physicians, the NHI’s payment was actually given to the hospital, which then in theory delivered payments to its own physicians who were treating P4P patients. However, we do not know whether the administrators of hospitals actually passed this P4P benefit on to physicians. Every hospital has its own physician fee policy. Second, some U.K. studies related to P4P exclusion also performed analysis at the practice level (Doran et al. 2008; Gravelle et al. 2010). Third, our physician-level data are incomplete, and we have records only of physician ids in our database. Because of these reasons, we decided to analyze exclusion behavior at the hospital level. In spite of this orientation, 65 percent of the variance in patient participation is explained by hospital characteristics, which indicates that these may factor significantly in patient enrollment in P4P.

Implications for Policy and Payment Reform

Based on our findings, we recommend that the government would benefit most from carrying out a deliberative and stepwise reform by first executing risk adjustment in the DM-P4P program and then gradually investing more money to cover the hospital costs of running the P4P program. Then, finally, the government will be in a well-grounded position, in terms of anticipating the reactions to the incentives and estimating its costs, to fully implement mandated participation in the P4P program.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The authors would like to thank the Bureau of National Health Insurance (No. DOH97-NH-1006) and Ministry of Education of Taiwan (No. 99RH0021) through its “Aiming for the Top University and Elite Research Center Development Plan” for their financial supports. In addition, the authors are also very grateful to Tung-Liang Chiang, Ph.D., for his support.

Disclosures: None.

Disclaimers: None.

REFERENCES

- Asch, S. M., E. A. Kerr, J. Keeseey, J. L. Adams, C. M. Setodji, S. Malik, and E. A. McGlynn. 2006. “Who Is at Greatest Risk for Receiving Poor-Quality Health Care?” *New England Journal of Medicine* 354 (11): 1147–56.
- Bardenheier, B. H., A. Shefer, L. Barker, C. A. Winston, and C. K. Sionean. 2005. “Public Health Application Comparing Multilevel Analysis with Logistic Regression: Immunization Coverage among Long-Term Care Facility Residents.” *Annals of Epidemiology* 15 (10): 749–55.
- Bokhour, B. G., J. F. Burgess Jr., J. M. Hook, B. White, D. Berlowitz, M. R. Guldin, M. Meterko, and G. J. Young. 2006. “Incentive Implementation in Physician Practices: A Qualitative Study of Practice Executive Perspectives on Pay for Performance.” *Medical Care and Research Review* 63 (suppl 1): 73S–95S.
- British Medical Association (BMA). 2009. “Quality and Outcomes Framework Guidance for GMS Contract 2009/10” [accessed on August 6, 2009]. Available at http://www.bma.org.uk/images/qof0309_tcm41-184025.pdf
- Centre for Studies in Social Sciences. 2009. “Mapping the Socially Excluded: Beyond Poverty Measurements” [accessed on July 4, 2010]. Available at http://www.cssscal.org/report_UNICEF-FEB09.pdf

- Chang, H. 2004. "Quality-Based Payment—Taiwan's Experience." In *Annual Research Meeting*, Academy Health.
- Chen, H. F., C. A. Ho, and C. Y. Li. 2006. "Age and Sex May Significantly Interact with Diabetes on the Risks of Lower-Extremity Amputation and Peripheral Revascularization Procedures: Evidence from a Cohort of a Half-Million Diabetic Patients." *Diabetes Care* 29 (11): 2409–14.
- Chiou, S. T., H. D. Lin, N. C. Yu, H. K. Hseuh, L. H. Lin, L. T. Lin, T. J. Chen, and M. S. Lai. 2001. "An Initial Assessment of the Feasibility and Effectiveness of Implementing Diabetes Shared Care System in Taiwan—Some Experiences from I-Lan County." *Diabetes Research and Clinical Practice* 54 (suppl 1): S67–73.
- Cho, S. H. 2003. "Using Multilevel Analysis in Patient and Organizational Outcomes Research." *Nursing Research* 52 (1): 61–5.
- Clarke, P. 2008. "When Can Group Level Clustering Be Ignored? Multilevel Models versus Single-Level Models with Sparse Data." *Journal of Epidemiology and Community Health* 62 (8): 752–8.
- Curcin, V., A. Bottle, M. Molokhia, C. Millett, and A. Majeed. 2010. "Towards a Scientific Workflow Methodology for Primary Care Database Studies." *Statistical Methods in Medical Research* 19 (4): 378–93.
- D'Errigo, P., M. E. Tosti, D. Fusco, C. A. Perucci, and F. Seccareccia. 2007. "Use of Hierarchical Models to Evaluate Performance of Cardiac Surgery Centres in the Italian CABG Outcome Study." *BMC Medical Research Methodology* 7: 29.
- David, A. H., and B. G. Mark. 1998. "Centering Decisions in Hierarchical Linear Models: Implications for Research in Organizations." *Journal of Management* 24 (5): 623.
- Doran, T., C. Fullwood, H. Gravelle, D. Reeves, E. Kontopantelis, U. Hiroeh, and M. Roland. 2006. "Pay-for-Performance Programs in Family Practices in the United Kingdom." *New England Journal of Medicine* 355 (4): 375–84.
- Doran, T., C. Fullwood, D. Reeves, H. Gravelle, and M. Roland. 2008. "Exclusion of Patients from Pay-for-Performance Targets by English Physicians." *New England Journal of Medicine* 359 (3): 274–84.
- Epstein, A. M. 2006. "Paying for Performance in the United States and Abroad." *New England Journal of Medicine* 355 (4): 406–8.
- Epstein, A. M., T. H. Lee, and M. B. Hamel. 2004. "Paying Physicians for High-Quality Care." *New England Journal of Medicine* 350 (4): 406–10.
- Fleetcroft, R., N. Steel, R. Cookson, and A. Howe. 2008. "'Mind the Gap!' Evaluation of the Performance Gap Attributable to Exception Reporting and Target Thresholds in the New GMS Contract: National Database Analysis." *BMC Health Services Research* 8: 131.
- Gravelle, H., M. Sutton, and A. Ma. 2010. "Doctor Behaviour under a Pay for Performance Contract: Treating, Cheating and Case Finding?" *Economic Journal* 120 (542): F129–56.
- Gulland, A. 2003a. "Negotiators Hit the Road to Explain the Proposed GP Contract." *British Medical Journal* 326: 519.
- . 2003b. "No Practice Will Lose out under New GP Contract Agreement Says BMA." *British Medical Journal* 323: 730.

- Halladay, J. R., S. C. Stearns, T. Wroth, L. Spragens, S. Hofstetter, S. Zimmerman, and P. D. Sloane. 2009. "Cost to Primary Care Practices of Responding to Payer Requests for Quality and Performance Data." *Annals of Family Medicine* 7 (6): 495–503.
- Lai, M. S., K. P. Chung, J. L. Hu, and T. T. Chen. 2009. *The Effect of Outcome-Based Diabetes Mellitus Pay-for-Performance Program in Taiwan*, National Health Insurance Technical Report. DOH98-NH-1003.
- Landon, B. E., S. L. Normand, D. Blumenthal, and J. Daley. 2003. "Physician Clinical Performance Assessment: Prospects and Barriers." *Journal of the American Medical Association* 290 (9): 1183–9.
- Lee, T. T., S. H. Cheng, C. C. Chen, and M. S. Lai. 2010. "A Pay-for-Performance Program for Diabetes Care in Taiwan: A Preliminary Assessment." *American Journal of Managed Care* 16 (1): 65–9.
- Lin, C. C., M. S. Lai, C. Y. Syu, S. C. Chang, and F. Y. Tseng. 2005. "Accuracy of Diabetes Diagnosis in Health Insurance Claims Data in Taiwan." *Journal of the Formosan Medical Association* 104 (3): 157–63.
- McGraw, K. O., and S.P. Wong. 1996. "Forming Inferences about Some Intraclass Correlation Coefficients." *Psychological Methods* 1 (1): 30–46.
- Meduru, P., D. Helmer, M. Rajan, C. L. Tseng, L. Pogach, and U. Sambamoorthi. 2007. "Chronic Illness with Complexity: Implications for Performance Measurement of Optimal Glycemic Control." *Journal of General Internal Medicine* 22 (suppl 3): 408–18.
- Merlo, J., B. Chaix, H. Ohlsson, A. Beckman, K. Johnell, P. Hjerpe, L. Rastam, and K. Larsen. 2006. "A Brief Conceptual Tutorial of Multilevel Analysis in Social Epidemiology: Using Measures of Clustering in Multilevel Logistic Regression to Investigate Contextual Phenomena." *Journal of Epidemiology and Community Health* 60 (4): 290–7.
- Pham, H. H., D. Schrag, A. S. O'Malley, B. Wu, and P. B. Bach. 2007. "Care Patterns in Medicare and Their Implications for Pay for Performance." *New England Journal of Medicine* 356 (11): 1130–9.
- Raudenbush, S. W., and A. S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage Publications.
- Reeves, D., S. M. Campbell, J. Adams, P. G. Shekelle, E. Kontopantelis, and M. O. Roland. 2007. "Combining Multiple Indicators of Clinical Quality: An Evaluation of Different Analytic Approaches." *Medical Care* 45 (6): 489–96.
- Rosenzweig, J. L., K. Weinger, L. Poirier-Solomon, and M. Rushton. 2002. "Use of a Disease Severity Index for Evaluation of Healthcare Costs and Management of Comorbidities of Patients with Diabetes Mellitus." *American Journal of Managed Care* 8 (11): 950–8.
- Ryan, A. M. 2009. "Has Pay-for-Performance Decreased Access for Minority Patients?" *Health Services Research* 45 (1): 6–23.
- Schootman, M., D. B. Jeff, W. E. Gillanders, Y. Yan, B. Jenkins, and R. Aft. 2007. "Geographic Clustering of Adequate Diagnostic Follow-Up after Abnormal Screening Results for Breast Cancer among Low-Income Women in Missouri." *Annals of Epidemiology* 17 (9): 704–12.

- Selby, J. V., A. J. Karter, L. M. Ackerson, A. Ferrara, and J. Liu. 2001. "Developing a Prediction Rule from Automated Clinical Databases to Identify High-Risk Patients in a Large Population with Diabetes." *Diabetes Care* 24 (9): 1547–55.
- Shen, Y. 2003. "Selection Incentives in a Performance-Based Contracting System." *Health Services Research* 38 (2): 535–52.
- Shwartz, M., J. Ren, E. A. Pekoz, X. Wang, A. B. Cohen, and J. D. Restuccia. 2008. "Estimating a Composite Measure of Hospital Quality from the Hospital Compare Database: Differences When Using a Bayesian Hierarchical Latent Variable Model versus Denominator-Based Weights." *Medical Care* 46 (8): 778–85.
- Sigfrid, L. A., C. Turner, D. Crook, and S. Ray. 2006. "Using the UK Primary Care Quality and Outcomes Framework to Audit Health Care Equity: Preliminary Data on Diabetes Management." *Journal of Public Health* 28 (3): 221–5.
- Slonim, A. D., J. P. Marcin, W. Turenne, M. Hall, and J. G. Joseph. 2007. "Pediatric Patient Safety Events during Hospitalization: Approaches to Accounting for Institution-Level Effects." *Health Services Research* 42 (6, Part 1): 2275–93.
- Tanenbaum, S. J. 2009. "Pay for Performance in Medicare: Evidentiary Irony and the Politics of Value." *Journal of Health Politics, Policy and Law* 34 (5): 717–46.
- Werner, R. M., and D. A. Asch. 2005. "The Unintended Consequences of Publicly Reporting Quality Information." *Journal of the American Medical Association* 293 (10): 1239–44.
- Young, B. A., E. Lin, M. Von Korff, G. Simon, P. Ciechanowski, E. J. Ludman, S. Everson-Stewart, L. Kinder, M. Oliver, E. J. Boyko, and W. J. Katon. 2008. "Diabetes Complications Severity Index and Risk of Mortality, Hospitalization, and Healthcare Utilization." *American Journal of Managed Care* 14 (1): 15–23.
- Young, G. J. 2008. "Can Multi-Level Research Help Us Design Pay-for-Performance Programs?" *Medical Care* 46 (2): 109–11.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Table SA1: Factors Associated with the Exclusion of DM Patients from P4P Programs.

Table SA2: Factors Associated with the Exclusion of DM Patients from P4P Programs Using a Logistic Model.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.